A Human Body Posture Identification Algorithm Based on Kinect

Tong Wang
School of mechatronics engineering
Harbin Institute of Technology
Harbin, China
Email:step186@163.com

Meng Wang
Chinese language and literature
HeiBei University
BaoDing, China
Email:18331066358@163.com

Xiaoliang Liu *
School of mechatronics engineering
Harbin Institute of Technology
Harbin, China
Email:lxlhit@foxmail.com
* Corresponding Author

Tongyao Wang
Visual Communication Design
Shandong Finance and Economics University
JiNan, China
Email:w_tong@qq.com

Abstract—This study was performed to improve the computer identify of the human body posture more accurately during human-computer interaction. In order to get a better algorithm, the author applies Kinect for Xbox device, Hausdorff Distance theory and Joint Angle measurement method to identify the human body posture. According to the actual situation, the author modifies the Hausdorff distance to mean Hausdorff distance; and sets one joint point as the reference point to improve the stability of angle measurement system. During the identification process, the Kinect gets the position of the human body joint points, and calculates the Euclidean distance between two joints, then uses cosine theorem to calculate the angle of two ligatures which are connecting three joints to define the body posture. The experimental results show that this algorithm can measure in time joint angles, and identify the body posture accurately. The study has an important reference meaning in human computer interaction.

Keywords-posture identification;Mean Hausdorff Distance;angle measurement; Kinect; human-computer interaction

I. INTRODUCTION

With the coming of Industrie 4.0, it is more and more important to realize Machine automation and intelligence. Machine's intelligence used to rely on computer control, and the natural and exact interaction between human and computer will greatly improve the working efficiency. Human-computer interaction relies on computer to implement the identification of specific action including posture, gestures, facial expressions[1]. Among the diverse technology of Human-computer interaction, the application of Kinect technology, especially the recognition of human body posture using Kinect, has become a research hotspot[2,3].

There are usually two ways to recognize human body posture:1 The direct method to test whether the human body consistent with human body posture model.2 A method based on forecast, the outline of a given characteristic method.

Due to the complexity of human movement and high visual rendering degeneration, the accurate identification of unmarked human body posture is extremely difficult[5,6,7]. This article focuses on how to improve the accuracy of human body gesture recognition, the method is based on the existing Kinect device technology and mean Hausdorff distance theory, proposed a new pattern of angle calculation.

II. KINECT STRUCTURE AND KEY TECHNOLOGY

A. *Kinect Structure*

Kinect is used for XBOX in the research. It is the first device series that Microsoft has launched, which does not need the handle control, it can directly complete human body gesture identification and equipment control, which is shown in Fig. 1. XBOX includes: a pair of 3D depth camera, a RGB camera, a set of microphone array and a rotating motors, it has the real-time dynamic capture function, voice transmission function, image transmission and multiplayer interactive functions[4,8].

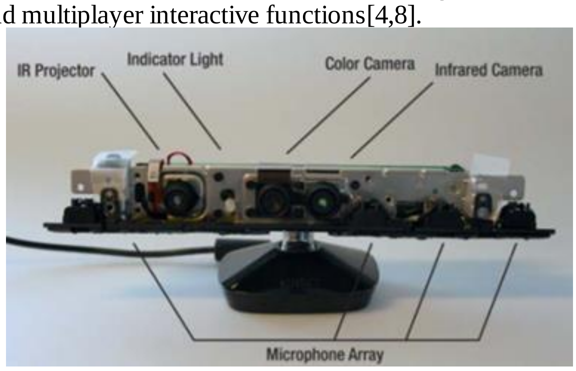


Figure 1. Kinect Structure

There is an infrared transmitter on the left side, a RGB camera in the middle and a Infrared ray CMOS camera. The cameras on the left and right side is used as 3D depth image sensor, mainly to retrieve the operators' action. The

RGB camera is used to identify the operators' identification[4,9]. On the bottom there is a motor, which can turn up and down to change the angle of the device. The Kinect's function is shown in the following table 1.

TABLE I. KINECT STRUCTURE AND FUNCTION

Structure	Function
The microphone	Collect sound signal
The motor	Adjust the camera pitching angle
Infrared transmitter	Infrared launch ,formation of speckle image
Infrared receiver	Collection of speckle image, the establishment of depth image
RGB camera	Identify the identity of the operator

B. Kinect Key Technology

Generating 3D data is the most important function of Kinect[10]. Its working principle is: the infrared camera obtain infrared data transfer to the Windows SDK toolkit, the computer generates the data then produces depth image data. In a certain depth image frames, each pixel contains information of distance is from the Kinect infrared transmitter to the captured point, the distance is calculated in millimeters[11].

Kinect can generate three different resolution depth images, a depth pixel consist 2 bytes. The first 13 data record the depth data, the last 3data record user ID, as shown in Fig. 1. If the depth of the location data is not available then the depth data is displayed as 0[12].

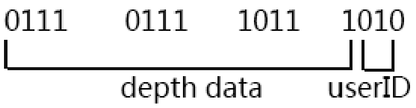


Figure 2. Structure of depth data in kinect

For Windows SDK, Kinect toolkit divide 3D information of human skeletal model is into 20 skeletal points. According to the 3D model of the human skeleton, Kinect for windows SDK can obtain real-time bone node position to calculate the angle between each bone node and the relative position[13]. If the data is got continuously, the movement vector of the skeleton node can be got. It can be identified that human postures and gestures are based on these data.

III. POSTURE IDENTIFICATION BASED ON MEAN AUSDORFF DISTANCE

There are various algorithms to identify the human body posture. In this paper, the Hausdorff distance algorithm is used[9], which modifies the Hausdorff distance to mean Hausdorff distance according to the actual situation. Two point at the space are given:

$$A=\{a_1,a_2,...,a_{m-1},a_m\},B=\{b_1,b_2,...,b_{n-1},b_n\},$$

The quantity is to evaluate the similarity is the Hausdorff distance, which is expressed as:

$$H(A,B)=\max\left[h(A,B),h(B,A)\right] \quad (1)$$

where:

$$h(A,B)=\max_{a\in A}\min_{b\in B}\|a-b\| \quad (2)$$

$$h(B,A)=\max_{a\in B}\min_{b\in A}\|b-a\| \quad (3)$$

$\|\bullet\|$ is defined as the distance norm from A to B at Euclidean distance; function $h(A,B)$ is the Directed Hausdorff distance from A to B, the value is proportional to the difference between set A and B.

In order to solve the large deviation caused by gusty noise by the algorithm of Hausdorff distance, Dubuisson summarizes the concept of part Hausdorff distance:

$$H_{LK}(A,B)=\max\left[h_L(A,B),h_K(B,A)\right] \quad (4)$$

where:

$$h_L(A,B)=\underset{a\in A}{\overset{th}{L}}d_B(a) \quad (5)$$

$$h_K(B,A)=\underset{a\in A}{\overset{th}{K}}d_A(b) \quad (6)$$

$\underset{a\in A}{\overset{th}{L}}$ --the Lth distance value is ranked from A to B, $0\le L\le m$;

$\underset{a\in A}{\overset{th}{K}}$ --the Kth distance value is ranked from B to A, $0\le K\le n$.

If the target is blocked or the part Hausdorff distance is mismatching caused by the rough noise in the matching image, the Hausdorff distance can be modifies as the Mean Hausdorff Distance (MHD):

$$H(A,B)=\max\left[h(A,B),h(B,A)\right] \quad (7)$$

where:

$$h(A,B)=\frac{1}{r}\sum_{a\in A}\min_{b\in B}\|a-b\| \quad (8)$$

$$h(B,A)=\frac{1}{s}\sum_{b\in B}\min_{a\in A}\|b-a\| \quad (9)$$

Where r and s are the number of elements in set A and B respectively. Then take use of the Mean Hausdorff Distance to evaluate the human body. Reduce the 5 sample sequence and pending recognition sequence to the five sequence $S_i(i=1,2,3,4,5)$ and pending recognition sequence S in 3-dimensions space. Then calculate the Mean Hausdorff Distance of $S_i(i=1,2,3,4,5)$ and S respectively ,the evaluation Criteria adopted is :

$$B=\arg\min_i(H(S,S_i)),(i=1,2,3,4,5) \quad (10)$$

IV. POSE RECOGNITION BASED ON JOINT POINT ANGLE MEASUREMENT

A. Establishment and transformation of coordinate

Before measuring the joint angle, establish the real world coordinate, setting the Kinect' sensor as the Coordinate origin, establish the depth image coordinate setting the origin of the depth image as the Coordinate origin. Then demarcate the Kinect: demarcate the infrared transmitter and the RGB camera.confirm the coordinate relationship between then depth image and RGB image:among the points in the 3D space, the initial value of d_r is defined as :

$$d = K\tan(H \cdot d_r + L) - O \quad (11)$$

Where d is the depth value of point,the unit is centimeter ;

$H = 3.5\times10^{-4} rad;$

$K = 12.36cm;$

$L = 1.18rad$;

$O = 3.7cm$.

Then render the depth data as different color according to the different distance from the target to the sensor.

Complicated the depth image, transfer the depth coordinate (x_d, y_d, z_d) to the real world coordinate (x_w, y_w, z_w) , the transformation function is :

$$\begin{cases} x_w = (x_d - \frac{w}{2}) \cdot (z_w + D') \cdot F \cdot (\frac{w}{h}) \\ y_w = (y_d - \frac{h}{2}) \cdot (z_w + D') \cdot F \\ z_w = d \end{cases} \quad (12)$$

$D' = -10, F = 0.0021,$ the resolution $w \times h$ of Kinect is 640×480.

B. The calculation of joint angle

Transfer any two points $X(x_1, x_2, x_3)$, $Y(y_1, y_2, y_3)$ In the space coordinate system to the real coordinate of the joint, apply the the Euclidean distance formula to calculate the distance between two adjacent joint:

$$D(X,Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2} \quad (13)$$

Take use of three joint points, calculate the joint angle: use the transformation function to calculate the real position coordinate of the joint, then calculate the distance of three adjacent points:

$$\begin{cases} a = D(B,C) \\ b = D(A,C) \\ c = D(A,B) \end{cases} \quad (14)$$

At last apply the Cosine theorem to calculate the joint angle:

$$\theta = \cos^{-1}\frac{(a^2 + c^2 - b^2)}{2ac} \quad (15)$$

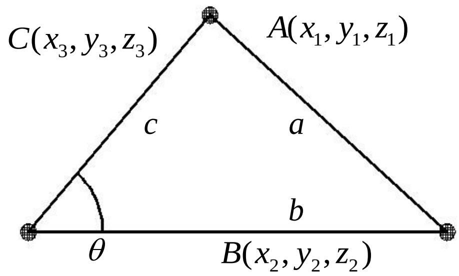


Figure 3. structure of joint angle

Analyze the above calculation process, the three joints in the space is relatively unstable, the calculated angle is not exact enough. In order to eliminate the unstable problem, firstly, make one joint as the reference point, then find another point, finally calculate the angle between the ligature of the two points and the x-axis of the reference point. This algorithm keeps the reference point and the specified point relatively stable to reduce the shake of different joints, it is more accurate to calculate the angle between two joints, what's more, by translating the reference point along the x-axis to the specified point, the calculation process can be simplified.

C. Determine the range of angle

Apply the above algorithm to ensure the formula of the joint angles:

$$P_A = \{P_1, P_2, \theta, \beta\} \quad (16)$$

Where,

P_1 is the reference center point :

θ is the angle between joint P_2 and x-axis :

β is the angle deviation threshold.

By setting different angle relations between joints, more posture can be defined, what's more, the angle deviation threshold β can be set according to the accuracy requirement. In this study, the angle of joints can be set as $\theta_i (i = 1,2,3,4)$: θ_1 =left shoulder–left elbow , θ_2 = left elbow-left wrist , θ_3 = right shoulder–right elbow, θ_4 = right elbow-right wrist,the angle deviation threshold β =10°, and the definition of body posture meet this condition : $\Delta = (\theta_1, \theta_2, \theta_3, \theta_4, \beta)$.

TABLE II. ANGLE CORRESPONDING TO EVERY POSTURE

Posture	θ_1	θ_2	θ_3	θ_4	β
Hands horizontal	180	180	0	0	10
Hands down	270	270	270	270	10
Hands up	180	90	0	90	10
Left up-right down	180	90	270	270	10
Left up-right horizontal	180	90	0	0	10
Right up-left down	270	270	0	90	10
Right up-left horizontal	180	180	0	90	10

Posture matching experiment

Setting the angle deviation threshold β a constant,establish the database of the postures .Capture human joints, then judge if the four angles meet the following formula:

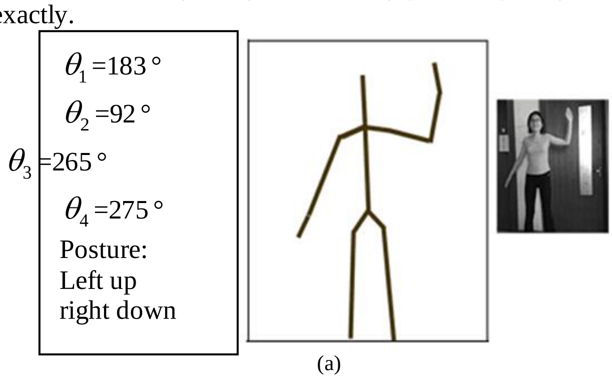
$$\max_{i\in N}\left|\theta_i - a_i\right| < \beta \qquad (17)$$

Where , θ_i is the actual measured angle , a_i is the preset angle .

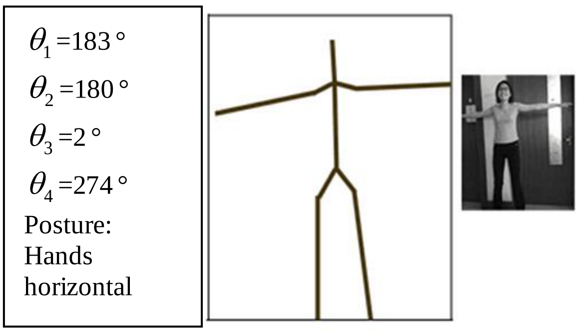
Take use of Kinect for XBOX, apply the new algorithm to identify the various body postures.

Experiment results: select four experiment results which are shown as Fig .4. (a-d) in the Fig .4 is Left hand up -right hand down, Hands horizontal, Hands down and the nonstandard Hands horizontal respectively. The posture' angle deviation in (a-c) is less than 10°, the monitor shows the posture's name; the posture' angle deviation in (d) exceeds the angle deviation threshold 10°, the monitor shows "failure".

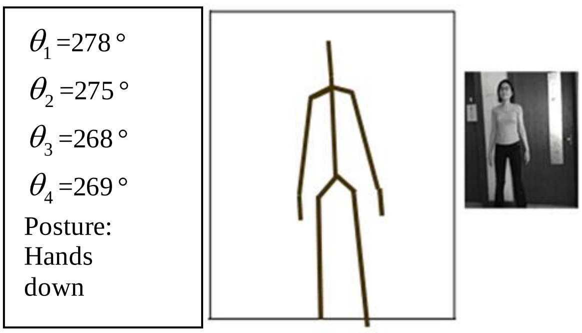
It can be concluded that: applying the algorithm, the Kinect can identity every human body posture quickly and exactly.



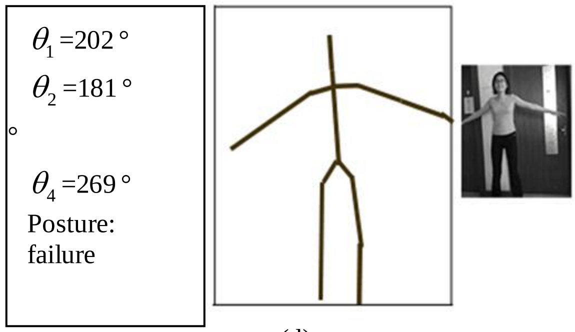
(a)



(b)



.(c)



.(d)

Figure 4. posture identification experiment

V. CONCLUSION

This paper introduced Kinect structure and the key technology, a new posture identification algorithm and method based on the modified mean Hausdorff distance theory, coordinate transformation angle measurement method. This method using Kinect, firstly, gets the position of joint points of the human body, by calculating the Euclidean distance between two joint points and using the cosine theorem to calculate the angle between the joint ligature the human body posture can be defined. Experimental results show that applying the algorithm, the Kinect can identity every human body posture quickly and exactly. This method has an important reference meaning in human computer interaction, especially for the posture identification.

REFERENCES

Chang Y J, Chen S F, Huang J D. A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities[J]. Research in developmental disabilities, 2011, 32(6): 2566-2570

Oikonomidis I, Kyriazis N, Argyros A A. Efficient model-based 3D tracking of hand articulations using Kinect[C]//BMVC. 2011, 1(2): 3..

Xia L, Chen C C, Aggarwal J K. Human detection using depth information by kinect[C]/Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on. IEEE, 2011: 15-22..

Henry P, Krainin M, Herbst E, et al. RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments[J]. The International Journal of Robotics Research, 2012, 31(5): 647-663..

Clark R A, Pua Y H, Fortin K, et al. Validity of the Microsoft Kinect for assessment of postural control[J]. Gait & posture, 2012, 36(3): 372-377..

Obdrzalek S, Kurillo G, Ofli F, et al. Accuracy and robustness of Kinect pose estimation in the context of coaching of elderly population[C]/Engineering in medicine and biology society (EMBC), 2012 annual international conference of the IEEE. IEEE, 2012: 1188-1193..

Janoch A, Karayev S, Jia Y, et al. A category-level 3d object dataset: Putting the kinect to work[M]/Consumer Depth Cameras for Computer Vision. Springer London, 2013: 141-165.

Boulos M N K, Blanchard B J, Walker C, et al. Web GIS in practice X: a Microsoft Kinect natural user interface for Google Earth navigation[J]. International journal of health geographics, 2011, 10(1): 45

[1] Villaroman N, Rowe D, Swan B. Teaching natural user interaction using OpenNI and the Microsoft Kinect sensor[C]//Proceedings of the 2011 conference on Information technology education. ACM, 2011: 227-232.:

[2] Stone E E, Skubic M. Passive in-home measurement of stride-to-stride gait variability comparing vision and Kinect sensing[C]/Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE. IEEE, 2011: 6491-6494.

[3] Alexiadis D S, Kelly P, Daras P, et al. Evaluating a dancer's performance using kinect-based skeleton tracking[C]/Proceedings of the 19th ACM international conference on Multimedia. ACM, 2011: 659-662.

[4] Nguyen C V, Izadi S, Lovell D. Modeling kinect sensor noise for improved 3d reconstruction and tracking[C]//3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on. IEEE, 2012: 524-530.

[5] [13] Li B Y L, Mian A S, Liu W, et al. Using kinect for face recognition under varying poses, expressions, illumination and disguise[C]/Applications of Computer Vision (WACV), 2013 IEEE Workshop on. IEEE, 2013: 186-192.